

Characterizing classes of regular languages using prefix codes of bounded synchronization delay

Volker Diekert and Tobias Walter*

University of Stuttgart, FMI
Universitätsstraße 38, 70569 Stuttgart, Germany
`{diekert,walter}@fmi.uni-stuttgart.de`

In this paper we continue a classical work of Schützenberger on codes with bounded synchronization delay. He was interested to characterize those regular languages where the groups in the syntactic monoid belong to a variety \mathbf{H} . He allowed operations on the language side which are union, intersection, concatenation and modified Kleene-star involving a mapping of a prefix code of bounded synchronization delay to a group $G \in \mathbf{H}$, but no complementation. In our notation this leads to the language classes $\text{SD}_G(A^\infty)$ and $\text{SD}_{\mathbf{H}}(A^\infty)$. Our main result shows that $\text{SD}_{\mathbf{H}}(A^\infty)$ always corresponds to the languages having syntactic monoids where all subgroups are in \mathbf{H} . Schützenberger showed this for a variety \mathbf{H} if \mathbf{H} contains Abelian groups, only. Our method shows the general result for all \mathbf{H} directly on finite and infinite words. Furthermore, we introduce the notion of *local Rees products* which refers to a simple type of classical Rees extensions. We give a decomposition of a monoid in terms of its groups and local Rees products. This gives a somewhat similar, but simpler decomposition than in Rhodes' synthesis theorem. Moreover, we need a singly exponential number of operations, only. Finally, our decomposition yields an answer to a question in a recent paper of Almeida and Klíma about varieties that are closed under Rees products.

1. Introduction

A fundamental result of Schützenberger characterizes the class of star-free languages SF as exactly those languages which are group-free, that is, aperiodic [13]. One usually

*Supported by the German Research Foundation (DFG) under grant DI 435/6-1.

abbreviates this result by $\text{SF} = \mathbf{Ap}$. Schützenberger also found another, but less prominent characterization of SF: the star-free languages are exactly the class of languages which can be defined inductively by finite languages and closure under union, concatenation, and the Kleene-star restricted to prefix codes of bounded synchronization delay [15]. This result is abbreviated by $\mathbf{Ap} = \text{SD}$. It is actually stronger than the famous $\text{SF} = \mathbf{Ap}$ because $\text{SD} \subseteq \text{SF} \subseteq \mathbf{Ap}$ is easy, so $\text{SF} = \mathbf{Ap}$ follows directly from $\mathbf{Ap} \subseteq \text{SD}$. The result $\mathbf{Ap} = \text{SD}$ has been extended to infinite words first in [4]. The extension to infinite words became possible thanks to a “local divisor approach”, which also is a main tool in this paper.

Schützenberger did not stop by showing $\mathbf{Ap} = \text{SD}$. In retrospective he started a program: in [14] he was able to prove an analogue of $\mathbf{Ap} = \text{SD}$ for languages where syntactic monoids have Abelian subgroups, only. In our notation $\mathbf{Ap} = \text{SD}$ means $\overline{\mathbf{I}}(A^\infty) = \text{SD}_1(A^\infty)$; and the main result in [14] is “essentially” equivalent to $\overline{\mathbf{Ab}}(A^*) = \text{SD}_{\mathbf{Ab}}(A^*)$. (We write “essentially” because using the structure theory of Abelian groups, a sharper version than $\overline{\mathbf{Ab}}(A^*) = \text{SD}_{\mathbf{Ab}}(A^*)$ is possible.) The proofs [14] use deep results in semigroup theory; and no such result beyond Abelian groups was known so far. Our result generalizes $\overline{\mathbf{Ab}}(A^\infty) = \text{SD}_{\mathbf{Ab}}(A^\infty)$ to every variety \mathbf{H} of finite groups: we show $\overline{\mathbf{H}}(A^\infty) = \text{SD}_{\mathbf{H}}(A^\infty)$. We were able to prove it with much less technical machinery compared to [14]. For example, no knowledge in Krohn-Rhodes theory is required.

Actually, our result is a generalization of $\overline{\mathbf{Ab}}(A^*) = \text{SD}_{\mathbf{Ab}}(A^*)$ [14] and also of $\mathbf{Ap}(A^\infty) = \text{SD}(A^\infty)$ [4]. More precisely, we give a characterization of languages which are recognized by monoids where all subgroups belong to \mathbf{H} . The characterization uses an inductive scheme starting with all finite subsets of finite words, allows concatenation, union, no(!) complementation, but a restricted use of a generalized Kleene-star (and ω -power in the case of infinite words). Let us explain the *generalized Kleene-star* in our context. Instead of putting the star above a single language, consider first a disjoint union $K = \bigcup \{K_g \mid g \in G\}$ where G is a finite group and each K_g is regular in A^* . The “generalized star” associates with such a disjoint union the following language:

$$\{u_{g_1} \cdots u_{g_k} \in K^* \mid u_{g_i} \in K_{g_i} \wedge g_1 \cdots g_k = 1 \in G\}.$$

Clearly, we obtain a regular language, but without any restriction, allowing such a “general star” yields all regular languages, even in the case of the trivial group. So, the construction is of no interest without a simultaneous restriction. The restriction considered in [14] yields an inductive scheme to define a class \mathcal{C} . The restriction says that such a generalized Kleene-star is allowed only over a disjoint union $K = \bigcup \{K_g \mid g \in G\}$ where each K_g already belongs to \mathcal{C} and where K is, in addition, a prefix code of bounded synchronization delay. The initials in “synchronization delay” led to the notation SD; and an indexed version SD_G (resp. $\text{SD}_{\mathbf{H}}$) refers to “synchronization delay over G ” (resp. over a finite group in \mathbf{H}). Since we also deal with infinite words we apply the same restriction to ω -powers.

Our results give also a new characterization for various other classes. For example, by a result of Straubing, Thérien and Thomas [18], the class of languages, having syntactic monoids where all subgroups are solvable, coincides with $(\text{FO} + \text{MOD})[<]$. Here, $(\text{FO} +$

$\text{MOD})[<]$ means the class of languages defined by the logic $(\text{FO} + \text{MOD})[<]$. Thus, we are able to give a new language characterization: $(\text{FO} + \text{MOD})[<](A^\infty) = \text{SD}_{\text{Sol}}(A^\infty)$.

Moreover, as a sort of byproduct of $\overline{\mathbf{H}} = \text{SD}_{\mathbf{H}}$, we obtain a simple and purely algebraic characterization of the monoids in $\overline{\mathbf{H}}$. Every monoid in $\overline{\mathbf{H}}$ can be decomposed in at most exponentially many iterated Rees products of groups in \mathbf{H} . The iteration uses only a very restricted version of Rees extensions: *local Rees products*. This means we obtain every finite monoid which is not a group as a divisor of a Rees extension between two proper divisors of M , one of them a proper submonoid, the other one a “local divisor”.

Our decomposition result is similar to the synthesis theory of Rhodes and Allen [11]. Moreover, our technique gives a singly exponential bound on the number of operations whereas no such bound was known by [11]. Finally, using this decomposition, we answer a recent question of Almeida and Klíma [1] concerning varieties which are closed under Rees products.

2. Preliminaries

Throughout, A denotes a finite alphabet and A^* is the free monoid over A . It consists of all finite words. The empty word is denoted by 1 as the neutral elements in other monoids or groups. The set of non-empty finite words is A^+ ; it is the free semigroup over A . By A^ω we denote the set of all infinite words with letters in A . For a set $K \subseteq A^*$, we let $K^\omega = \{u_1 u_2 \cdots \mid u_i \in K \text{ non-empty}, i \in \mathbb{N}\} \subseteq A^\omega$. In particular, $K^\omega = (K \setminus \{1\})^\omega$. Since our results concern finite and infinite words, it is convenient to treat finite and infinite words simultaneously. We define $A^\infty = A^* \cup A^\omega$ to be the set of finite or infinite words. Accordingly, a *language* L is a subset of A^∞ . We say that L is *regular*, if first, $L \cap A^*$ is regular and second, $L \cap A^\omega$ is ω -regular in the standard meaning of formal language theory. In order to study regular languages algebraically, one considers finite monoids. A *divisor* of a monoid M is a monoid N which is a homomorphic image of a subsemigroup of M . In this case we write $N \preceq M$. A subsemigroup S of M is in our setting a divisor if and only if S is a monoid (but not necessarily a submonoid of M). A *variety* of finite monoids – hence, in Birkhoff’s setting: a *pseudovariety* – is a class of finite monoids \mathbf{V} which is closed under finite direct products and under division:

- If I is a finite index set and $M_i \in \mathbf{V}$ for each $i \in I$, then $\prod_{i \in I} M_i \in \mathbf{V}$. In particular, the trivial group $\{1\}$ belongs to \mathbf{V} .
- If $M \in \mathbf{V}$ and $N \preceq M$, then $N \in \mathbf{V}$.

Classical formal language theory states “regular” is the same as “recognizable”. This means: $L \subseteq A^*$ is regular if and only if its syntactic monoid is finite; $L \subseteq A^\omega$ is regular if and only if its syntactic monoid (in the sense of Arnold) is finite and, in addition, L is saturated by the syntactic congruence, see eg. [9, 19]. Here we use a notion of recognizability which applies to languages $L \subseteq A^\infty$. Let $\varphi : A^* \rightarrow M$ be a homomorphism to a finite monoid M . First, we define a relation \sim_φ as follows. If $u \in A^*$ is a finite word, then we write $u \sim_\varphi v$ if v is finite and $\varphi(u) = \varphi(v)$. If $u \in A^\omega$ is an infinite word, then we write $u \sim_\varphi v$ if v is infinite and if there are factorizations

$u = u_1 u_2 \dots$ and $v = v_1 v_2 \dots$ into finite nonempty words such that $\varphi(u_i) = \varphi(v_i)$ for all $i \geq 1$. It is easy to see that \sim_φ is not transitive on infinite words, in general. Therefore, we consider its transitive closure \approx_φ . If $u, v \in A^*$, then we have

$$u \sim_\varphi v \iff u \approx_\varphi v \iff \varphi(u) = \varphi(v).$$

If $\alpha, \beta \in A^\omega$, then we have $\alpha \approx_\varphi \beta$ if and only if there is sequence of infinite words $\alpha_0, \dots, \alpha_k$ such that

$$\alpha = \alpha_0 \sim_\varphi \dots \sim_\varphi \alpha_k = \beta.$$

We say that $L \subseteq A^\infty$ is *recognizable* by M if there exists a homomorphism $\varphi : A^* \rightarrow M$ such that $u \in L$ and $u \sim_\varphi v$ implies $v \in L$. We also say that M or φ recognizes L in this case.

The connection to the classical notation is as follows. A regular language $L \subseteq A^\infty$ is recognizable (in our sense) by φ if and only if the syntactic monoids of $L \cap A^*$ and $L \cap A^\omega$ are divisors of M (in the classical sense).

Every variety \mathbf{V} defines a family of regular languages $\mathbf{V}(A^\infty)$ as follows: we let $L \in \mathbf{V}(A^\infty)$ if there exists a monoid $M \in \mathbf{V}$ which recognizes L . Further, we define $\mathbf{V}(A^*) = \{L \subseteq A^* \mid L \in \mathbf{V}(A^\infty)\}$ and $\mathbf{V}(A^\omega) = \{L \subseteq A^\omega \mid L \in \mathbf{V}(A^\infty)\}$. A variety of finite groups is a variety of finite monoids which contains only groups. Throughout \mathbf{H} denotes a variety of finite groups. Special cases are the varieties

- **1**: the trivial group $\{1\}$, only.
- **Ab**: all finite Abelian groups.
- **Sol**: all finite solvable groups.
- **Sol_q**: all finite solvable groups where the order is divisible by some power of q .
- **G**: all finite groups.

According to standard notation $\overline{\mathbf{H}}$ denotes the variety of finite monoids where all subgroups belong to \mathbf{H} . It is not completely obvious, but a classical fact [8], that $\overline{\mathbf{H}}$ is indeed a variety. In fact, it is the maximal variety \mathbf{V} such that $\mathbf{V} \cap \mathbf{G} = \mathbf{H}$.

Clearly, $\overline{\mathbf{G}}$ is the class of all finite monoids. The most prominent subclass is $\overline{\mathbf{1}}$: it is the variety of aperiodic monoids **Ap**. The class $\mathbf{Ap}(A^\infty) = \overline{\mathbf{1}}(A^\infty)$ admits various other characterizations as subsets of A^∞ . For example, it is the class of star-free languages $\mathbf{SF}(A^\infty)$, it is the class of first-order definable languages, and it is the class of definable languages in linear temporal logic over finite or infinite words: $\mathbf{LTL}(A^\infty)$.

Local divisors. Let M be a finite monoid and $c \in M$. Consider the set $cM \cap Mc$ with a new multiplication \circ which is defined as follows:

$$mc \circ cn = mc n.$$

A straightforward calculation shows that $cM \cap Mc$ becomes a monoid with this operation where the neutral element of M_c is c . Thus, the structure $M_c = (cM \cap Mc, \circ, c)$ defines

a monoid. We say that M_c is the *local divisor* of M at c . If c is a unit, then M_c is isomorphic to M . If $c = c^2$, then M_c is the standard “local monoid” at the idempotent c .

The important fact is that M_c is always a divisor of M and that $|M_c| < |M|$ as soon as c is not a unit of M . Indeed, the mapping $\lambda_c : \{x \in M \mid cx \in M_c\} \rightarrow M_c$ given by $\lambda_c(x) = cx$ is a surjective homomorphism. Moreover, if c is not a unit, then $1 \notin cM \cap M_c$, hence $|M_c| < |M|$. Thus, if M belongs to some variety \mathbf{V} , then M_c belongs to the same variety. If M is not a group, then we find some nonunit $c \in M$ and the local divisor M_c is smaller than M . This makes the construction useful for induction. For a survey on the local divisor technique we refer to [5].

Rees extensions. Let N, L be monoids and $\rho : N \rightarrow L$ be any mapping. The *Rees extension* $\text{Rees}(N, L, \rho)$ is a classical construction for monoids [10, 12], frequently described in terms of matrices. Here, we use an equivalent definition as in [6]. As a set we define

$$\text{Rees}(N, L, \rho) = N \cup N \times L \times N.$$

The multiplication \cdot on $\text{Rees}(N, L, \rho)$ is given by

$$\begin{aligned} n \cdot n' &= nn' && \text{for } n, n' \in N, \\ n \cdot (n_1, m, n_2) \cdot n' &= (nn_1, m, n_2n') && \text{for } n, n', n_1, n_2 \in N, m \in L, \\ (n_1, m, n_2) \cdot (n'_1, m', n'_2) &= (n_1, m\rho(n_2n'_1)m', n'_2) && \text{for } n_1, n'_1, n_2, n'_2 \in N, m, m' \in L. \end{aligned}$$

The neutral element of $\text{Rees}(N, L, \rho)$ is $1 \in N$ and $N \subseteq \text{Rees}(N, L, \rho)$ is an embedding of monoids. In general, L is not a divisor of $\text{Rees}(N, L, \rho)$. The following property holds.

Lemma 1. *Let $N \preceq N'$ and $L \preceq L'$. Given $\rho : N \rightarrow L$, there exists a mapping $\rho' : N' \rightarrow L'$ such that $\text{Rees}(N, L, \rho)$ is a divisor of $\text{Rees}(N', L', \rho')$.*

Proof. First, assume that N (resp. L) is submonoid in N' (resp. L'). Let $\rho' : N' \rightarrow L'$ be any function such that $\rho'|_N = \rho$. The mapping $\pi : \text{Rees}(N, L, \rho) \rightarrow \text{Rees}(N', L', \rho')$ given by $\pi(n) = n$ and $\pi(n_1, \ell, n_2) = (n_1, \ell, n_2)$ is an injective homomorphism.

Second, let $\varphi : N' \rightarrow N$ and $\psi : L' \rightarrow L$ be surjective homomorphisms. Let $\rho' : N' \rightarrow L'$ be a function such that $\rho'(n) \in \psi^{-1}(\rho(\varphi(n)))$. Let $\pi : \text{Rees}(N', L', \rho') \rightarrow \text{Rees}(N, L, \rho)$ be the mapping defined by $\pi(n) = \varphi(n)$ and $\pi(n_1, \ell, n_2) = (\varphi(n_1), \psi(\ell), \varphi(n_2))$. It is clear that π is surjective. It is a homomorphism since

$$\begin{aligned} \pi((n_1, \ell, n_2) \cdot (n'_1, \ell', n'_2)) &= \pi(n_1, \ell\rho'(n_2n'_1)\ell', n'_2) = (\varphi(n_1), \psi(\ell) \underbrace{\psi(\rho'(n_2n'_1))}_{=\rho(\varphi(n_2n'_1))}, \varphi(n'_2)) \\ &= (\varphi(n_1), \psi(\ell), \varphi(n_2)) \cdot (\varphi(n'_1), \psi(\ell'), \varphi(n'_2)) = \pi(n_1, \ell, n_2) \cdot \pi(n'_1, \ell', n'_2). \end{aligned}$$

The result follows because \preceq is transitive. \square

We are mainly interested in the case where N and L are proper divisors of a given finite monoid M . This leads to the notion of local Rees monoids. More precisely, let M be a finite monoid, N by a proper submonoid of M and M_c be a local divisor of M at

c where c is not a unit. The *local Rees product* $\text{LocRees}(\mathbf{N}, \mathbf{M}_c)$ is defined as the Rees extension $\text{Rees}(\mathbf{N}, \mathbf{M}_c, \rho_c)$ where ρ_c denotes the mapping $\rho_c : N \rightarrow M_c; x \mapsto cxc$.

For a variety \mathbf{V} we define $\text{Rees}(\mathbf{V})$ to be the least variety which contains \mathbf{V} and is closed under taking Rees products and $\text{LocRees}(\mathbf{V})$ to be the least variety which contains \mathbf{V} and is closed under local Rees products.

2.1. Schützenberger's SD classes

Schützenberger gave a language theoretical characterization of the class of star-free languages $\text{SF}(A^*)$ avoiding complementation, but allowing the star-operation to prefix codes of bounded synchronization delay [15].

A language $K \subseteq A^+$ is called *prefix code* if it is *prefix-free*. That is: $u, uv \in K$ implies $u = uv$. A prefix-free language K is a code since every word $u \in K^*$ admits a unique factorization $u = u_1 \cdots u_k$ with $k \geq 0$ and $u_i \in K$. Note that the empty set \emptyset is considered to be a prefix code. More generally, if $L \subseteq A^+$ is any subset, then $K = L \setminus LA^+$ is a prefix code. A prefix code K has *bounded synchronization delay* if for some $d \in \mathbb{N}$ and for all $u, v, w \in A^*$ we have: if $uvw \in K^*$ and $v \in K^d$, then $uv \in K^*$. Note that the condition implies that for all $uvw \in K^*$ with $v \in K^d$, we have $w \in K^*$, too. If d is given explicitly, K has said to have synchronization delay d . Every subset $B \subseteq A$ (including the empty set) yields a prefix code with synchronization delay 0. If we have $c \in A \setminus B$, then B^*c is a prefix code with synchronization delay 1. If K is any prefix code with (or without) bounded synchronization delay, then K^m is a prefix code for all $m \in \mathbb{N}$, but for $m \geq 2$ it is never of bounded synchronization delay.

Let G be a finite group. By $\text{SD}_G(A^\infty)$ we denote the set of regular languages which is inductively defined as follows.

1. We let $\emptyset \in \text{SD}_G(A^\infty)$ and $\{a\} \in \text{SD}_G(A^\infty)$ for all letters $a \in A$.
2. If $L, K \in \text{SD}_G(A^\infty)$, then $L \cup K$ and $(L \cap A^*) \cdot K$ are both in $\text{SD}_G(A^\infty)$.
3. Let $K \subseteq A^+$ be a prefix code of bounded synchronization delay and $\gamma_K : K \rightarrow G$ be any mapping of K to the group G such that $\gamma_K^{-1}(g) \in \text{SD}_G(A^\infty)$ for all $g \in G$. We let $\gamma^{-1}(1) \in \text{SD}_G(A^\infty)$ and $\gamma^{-1}(1)^\omega \in \text{SD}_G(A^\infty)$, where $\gamma : K^* \rightarrow G$ denotes the canonical extension of γ_K to a homomorphism from the free submonoid $K^* \subseteq A^*$ to G .

We also define

$$\text{SD}_G(A^*) = \{L \subseteq A^* \mid L \in \text{SD}_G(A^\infty)\} \quad \text{and} \quad \text{SD}_G(A^\omega) = \{L \subseteq A^\omega \mid L \in \text{SD}_G(A^\infty)\}.$$

Note that for every homomorphism $\gamma : A^* \rightarrow G$ we have $\gamma^{-1}(1) \in \text{SD}_G(A^*)$ and $\gamma^{-1}(1)^\omega \in \text{SD}_G(A^\omega)$. This follows because first, A is a prefix code of bounded synchronization delay and second, all finite subsets of A are in $\text{SD}_G(A^*)$.

Unlike the case of star-free sets, the inductive definition of $\text{SD}_G(A^\infty)$ does not use any complementation. By induction: for $L \subseteq A^\infty$ we have $L \in \text{SD}_G(A^\infty)$ if and only if we can write $L = L_1 \cup L_2$ with $L_1 \in \text{SD}_G(A^*)$ and $L_2 \in \text{SD}_G(A^\omega)$. In the special case where

$G = \{1\}$ is the trivial group, we also simply write SD instead of $\text{SD}_{\{1\}}$. In this case the third condition can be rephrased in simpler terms as follows.

- If $K \in \text{SD}(A^*)$ is a prefix code of bounded synchronization delay, then $K^* \in \text{SD}(A^*)$ and $K^\omega \in \text{SD}(A^\omega)$.

In [14] Schützenberger showed (using a different notation) $\text{SD}_{\mathbf{H}}(A^*) \subseteq \overline{\mathbf{H}}(A^*)$, but the converse only for $\mathbf{H} \subseteq \mathbf{Ab}$, see Proposition 6 for the first inclusion. Our aim is to show $\overline{\mathbf{H}}(A^\infty) \subseteq \text{SD}_{\mathbf{H}}(A^*)$ for all \mathbf{H} , cf. Theorem 4. We begin with a technical lemma.

Lemma 2. *Let $K \subseteq A^+$ be a prefix code of bounded synchronization delay and let $\gamma : K^* \rightarrow G$ be a homomorphism such that $\gamma^{-1}(g) \cap K \in \text{SD}_G(A^*)$ for all $g \in G$, then we have $\gamma^{-1}(g) \in \text{SD}_G(A^*)$ for all $g \in G$.*

Proof. For each $w \in K^*$ we construct a language $L(w) \in \text{SD}_G(A^*)$ such that

- $w \in L(w) \subseteq \gamma^{-1}(\gamma(w))$,
- $|\{L(w) \mid w \in K^*\}| < \infty$.

Consider $w = u_1 \cdots u_k \in \gamma^{-1}(g)$ with $u_i \in K$. Define $P(w) = \{\gamma(u_1 \cdots u_i) \mid 1 \leq i \leq k\} \subseteq G$ to be the set of prefixes of w in G . We perform an induction on $|P(w)|$. The case $|P(w)| = 0$ implies $g = 1$. Hence, we let $L(w) = \gamma^{-1}(1)$; and we have $\gamma^{-1}(1) \in \text{SD}_G(A^*)$ by definition. Hence, we may assume $|P(w)| \geq 1$. Let $g_1 = \gamma(u_1)$ and choose i maximal such that $g_1 = \gamma(u_1 \cdots u_i)$. Then we have $u_1 \cdots u_i \in (K \cap \gamma^{-1}(g_1)) \cdot \gamma^{-1}(1)$. Define $w' = u_{i+1} \cdots u_k$. By maximality of i we have $|\{\gamma(u_1 \cdots u_j) \mid i < j \leq k\}| < |P(w)|$ because $P(w') = g_1^{-1} \cdot \{\gamma(u_1 \cdots u_j) \mid i < j \leq k\}$. By induction there exists $L(w')$ (and only a finite number of them); and we let $L(w) = (K \cap \gamma^{-1}(g_1)) \cdot \gamma^{-1}(1) \cdot L(w')$. The result follows because we can write $\gamma^{-1}(g) = \bigcup \{L(w) \mid w \in \gamma^{-1}(g)\}$ and this is a finite union. \square

Clearly, we have for all G : if $K \in \text{SD}_G(A^*)$ is a prefix code of bounded synchronization delay, then K^* and K^ω are both in $\text{SD}_G(A^\infty)$. As a special case, using the prefix code $K = \emptyset$, it holds $K^* = \{1\} \in \text{SD}_G(A^\infty)$. More generally, every finite language is in $\text{SD}_G(A^\infty)$. Note also that for $G' \leq G$ we have $\text{SD}_{G'}(A^\infty) \subseteq \text{SD}_G(A^\infty)$. In particular, $\bigcup \{\text{SD}_{G_i}(A^\infty) \mid i \in I\} \subseteq \text{SD}_{\prod_{i \in I} G_i}(A^\infty)$ for every finite index set I . This inclusion holds for every divisor of G as observed by the next lemma.

Lemma 3. $\text{SD}_H(A^\infty) \subseteq \text{SD}_G(A^\infty)$ holds for $H \preceq G$.

Proof. Inductively, it suffices to prove that $\gamma^{-1}(1), \gamma^{-1}(1)^\omega \in \text{SD}_G(A^\infty)$ for a prefix code $K \subseteq A^+$ of bounded synchronization delay and $\gamma : K^* \rightarrow H$ a homomorphism of the free monoid K^* to the group H such that $K \cap \gamma^{-1}(h) \in \text{SD}_G(A^\infty)$ for all $h \in H$. Without loss of generality we may assume that there exists a surjective homomorphism $\pi : G \rightarrow H$. Let $g_h \in G$ be elements such that $\pi(g_h) = h$. Let $\psi : K^* \rightarrow G$ be the homomorphism such that $\psi(u) = g_{\gamma(u)}$ for $u \in K$. By definition it holds $\gamma = \pi \circ \psi$. Now $K \cap \psi^{-1}(g_h) = K \cap \gamma^{-1}(h) \in \text{SD}_G(A^\infty)$ and $K \cap \psi^{-1}(g) = \emptyset$ if $g \neq g_h$ for all $h \in H$.

Thus, $\psi^{-1}(1), \psi^{-1}(1)^\omega \in \text{SD}_G(A^\infty)$ and by Lemma 2 we also have $\psi^{-1}(g) \in \text{SD}_G(A^\infty)$ for all $g \in G$. Note that

$$\begin{aligned}\gamma^{-1}(1) &= \bigcup_{\pi(g)=1} \psi^{-1}(g) \quad \text{and} \\ \gamma^{-1}(1)^\omega &= \bigcup_{\pi(g)=1} \psi^{-1}(g)\psi^{-1}(1)^\omega\end{aligned}$$

which proves that $\gamma^{-1}(1), \gamma^{-1}(1)^\omega \in \text{SD}_G(A^\infty)$. \square

We will formulate our results on the language classes $\text{SD}_G(A^\infty)$ to obtain finer results, however our main result then is formulated with the language class

$$\text{SD}_{\mathbf{H}}(A^\infty) = \bigcup \{ \text{SD}_G(A^\infty) \mid G \in \mathbf{H} \}.$$

The main result is the following equality between $\text{SD}_{\mathbf{H}}, \overline{\mathbf{H}}$ and $\text{LocRees}(\mathbf{H})$.

Theorem 4. *Let $L \subseteq A^\infty$ be a regular language and \mathbf{H} a variety of finite groups. Then the following properties are equivalent:*

1. $L \in \text{SD}_{\mathbf{H}}(A^\infty)$.
2. $L \in \overline{\mathbf{H}}(A^\infty)$.
3. $L \in \text{LocRees}(\mathbf{H})(A^\infty)$.

Corollary 5. *$\text{SD}_{\mathbf{H}}(A^\infty)$ is closed under complementation and intersection for every variety \mathbf{H} of finite groups.*

Proof. By Theorem 4 we have $\text{SD}_{\mathbf{H}}(A^\infty) = \overline{\mathbf{H}}(A^\infty)$ and $\overline{\mathbf{H}}(A^\infty)$ is closed under complementation and intersection. \square

The proof of Theorem 4 covers the next three sections.

3. Closure properties of $\text{SD}_{\mathbf{H}}$

In this section we prove the direction $1 \implies 2$ of Theorem 4. Therefore one has to study the closure properties under the operations given in the definition of $\text{SD}_{\mathbf{H}}(A^\infty)$, that is, one has to show that those operations do not introduce new groups.

The following proposition of Schützenberger shows that the operation $\gamma^{-1}(1)$ does not introduce new groups.

Proposition 6 ([14]). *Let $K \subseteq A^+$ be a prefix code of bounded synchronization delay and $\gamma_K : K \rightarrow G$ be a mapping such that $K_g = \gamma_K^{-1}(g)$ are regular languages for $g \in G$. Let $\gamma : K^* \rightarrow G$ be the homomorphism from the free submonoid K^* of A^* to the group G such that $\gamma|_K = \gamma_K$. View $\gamma^{-1}(1)$ as a subset of A^* . Then, subgroups in the syntactic monoid of the language $\gamma^{-1}(1)$ are either divisors of G or of the direct product $\prod_{g \in G} \text{Synt}(K_g)$.*

We will prove the same for $\gamma^{-1}(1)^\omega$, relying on Proposition 6 as a blackbox result. The concept used for transferring the properties to infinite words are Birget-Rhodes expansions [2, 3]. The Birget-Rhodes expansion of a monoid M is the monoid $\text{Exp}(M) = \{(X, m) \mid 1, m \in X \subseteq M\}$. The multiplication on $\text{Exp}(M)$ is given as a “semi-direct product”: $(X, m) \cdot (Y, n) = (X \cup m \cdot Y, m \cdot n)$. Note that M is isomorphic to the submonoid $\{(M, m) \mid m \in M\}$ of $\text{Exp}(M)$, that is, M is a divisor of $\text{Exp}(M)$. Moreover, the following lemma shows that the Birget-Rhodes expansion has the same groups as M .

Lemma 7. *Every group contained in $\text{Exp}(M)$ is isomorphic to some group in M .*

Proof. Let $G \subseteq \text{Exp}(M)$ be a group contained in $\text{Exp}(M)$ and let $(X, e) \in G$ be the unit in G . For every element $(Y, m) \in G$ we have $(X, e)(Y, m) = (X \cup eY, em) = (Y, m)$ and thus $X \subseteq Y$. Furthermore, $(Y, m)^{|G|} = (Y \cup \dots, e) = (X, e)$ and we conclude $X = Y$. Thus, $(X, m) \mapsto m$ is an injective embedding of G in M . \square

The idea behind the Birget-Rhodes expansion is that it stores the seen prefixes in a set. More formally, the following lemma holds.

Lemma 8. *Let $\varphi : A^* \rightarrow M$ be a homomorphism and $\psi : A^* \rightarrow \text{Exp}(M)$ be the homomorphism given by $\psi(a) = (\{1, \varphi(a)\}, \varphi(a))$. Let $u \in A^*$ and $\psi(u) = (X, \varphi(u))$. For every $m \in X$ there exists a prefix v of u such that $\varphi(v) = m$.*

Proof. We will prove this inductively. The statement is true if u is the empty word. Thus, consider $u = va$ for some letter $a \in A$. Let $\psi(v) = (Y, \varphi(v))$, then

$$\psi(u) = \psi(v) \cdot (\{1, \varphi(a)\}, \varphi(a)) = (Y \cup \{\varphi(v), \varphi(v)\varphi(a)\}, \varphi(u)).$$

Inductively, we obtain prefixes of v , and therefore also prefixes of u , for all elements of Y . The only (potentially) new element in X is $\varphi(u)$. This proves the claim. \square

A special kind of ω -regular languages are *arrow languages*. Let $L \subseteq A^*$ be a language. We define $\vec{L} = \{\alpha \in A^\omega \mid \text{infinitely many prefixes of } \alpha \text{ are in } L\}$ to be the arrow language of L . The set of arrow languages is exactly the set of deterministic languages [19]. The Birget-Rhodes expansion can be used to obtain a recognizing monoid for \vec{L} , given a monoid for L .

Proposition 9. *Let $L \subseteq A^*$ be some regular language and $\varphi : A^* \rightarrow M$ be a homomorphism which recognizes L , then \vec{L} is recognized by $\text{Exp}(M)$.*

Proof. Let $\psi : A^* \rightarrow \text{Exp}(M)$ be the homomorphism given by $\psi(a) = (\{1, \varphi(a)\}, \varphi(a))$. Let $\alpha \in \vec{L}$ and $\alpha \sim_\psi \beta$. We show that $\beta \in \vec{L}$. Let $\alpha = u_1 u_2 \dots$ and $\beta = v_1 v_2 \dots$ be factorizations such that $\psi(u_i) = \psi(v_i)$. Since $\alpha \in \vec{L}$, we may assume that for every i there exists a decomposition $u_i = u'_i u''_i$ such that $u_1 \dots u_{i-1} u'_i \in L$. By $\psi(u_i) = \psi(v_i)$ and Lemma 8, there exists a decomposition $v_i = v'_i v''_i$ such that $\varphi(u'_i) = \varphi(v'_i)$. Thus, $u_1 \dots u_{i-1} u'_i \sim_\varphi v_1 \dots v_{i-1} v'_i$ and therefore $v_1 \dots v_{i-1} v'_i \in L$. This implies $\beta \in \vec{L}$. \square

We are now ready to show the main result of this section, that is, every language in $\text{SD}_G(A^\infty)$ has only groups which are divisors of direct products of G . In particular, this implies $\text{SD}_{\mathbf{H}}(A^\infty) \subseteq \overline{\mathbf{H}}(A^\infty)$.

Proposition 10. *If $L \in \text{SD}_G(A^\infty)$, then all subgroups in $\text{Synt}(L)$ are a divisor of a direct product of copies of G .*

Proof. We will prove this inductively on the definition of $\text{SD}_G(A^\infty)$. The cases $\emptyset \in \text{SD}_G(A^\infty)$ and $\{a\} \in \text{SD}_G(A^\infty)$ for all letters $a \in A$ are straightforward, as they are recognized by aperiodic monoids. Let L, K be languages, such that their syntactic monoids contain only groups which are divisors of a direct product of G . The language $L \cup K$ is recognized by the direct product of their syntactic monoids which implies the statement. $(L \cap A^*) \cdot K$ is recognized by the Schützenberger product of their syntactic homomorphisms [7, Proposition 11.7.10]. The Schützenberger product does not introduce new groups [13]¹.

Let $K \subseteq A^+$ be a prefix code of bounded synchronization delay and $\gamma : K^* \rightarrow G$ be a homomorphism of the free monoid K^* to the group G such that for all $g \in G$ every subgroup of $\text{Synt}(K \cap \gamma^{-1}(g))$ is a divisor of a direct product of copies of G . Proposition 6 implies that every subgroup of $\text{Synt}(\gamma^{-1}(1))$ is a divisor of a direct product of copies of G . Note that $\gamma^{-1}(1)^\omega = \overrightarrow{\gamma^{-1}(1)}$ and therefore Proposition 9 and Lemma 7 imply that every subgroup of $\text{Synt}(\gamma^{-1}(1)^\omega)$ is a divisor of a direct product of copies of G . \square

4. The inclusion $\overline{\mathbf{H}}(A^\infty) \subseteq \text{SD}_{\mathbf{H}}(A^\infty)$

In this section we prove the direction $2 \implies 1$. We prove that if every subgroup of M is a divisor of G , then every language recognized by M is contained in $\text{SD}_G(A^\infty)$. This result is again finer than just the inequality $\overline{\mathbf{H}}(A^\infty) \subseteq \text{SD}_{\mathbf{H}}(A^\infty)$. The proof works by induction on $|M|$ and on the alphabet and decomposes every \approx_φ -class into several sets in $\text{SD}_G(A^\infty)$.

Proposition 11. *Let $L \subseteq A^\infty$ be recognized by $\varphi : A^* \rightarrow M$ and let G be a group such that every subgroup of M is a divisor of G , then $L \in \text{SD}_G(A^\infty)$. Moreover, L can be written as finite union*

$$L = L_0 \cup \bigcup_{i=1}^m L_i \cdot \gamma_i^{-1}(1)^\omega$$

for $L_i \in \text{SD}_G(A^*)$ and $\gamma_i : K_i^* \rightarrow G$ for prefix codes $K_i \in \text{SD}_G(A^*)$ of bounded synchronization delay with $\gamma_i^{-1}(g) \cap K_i \in \text{SD}_G(A^*)$ for all $g \in G$. All products in the expressions of L_i are unambiguous.

Proof. Let $\llbracket w \rrbracket_\varphi = \{v \in A^\infty \mid w \approx_\varphi v\}$ be the equivalence class of w . Since L is recognized by φ , it holds $L = \cup_{w \in L} \llbracket w \rrbracket_\varphi$. Our goal is to construct languages $L(w) \in \text{SD}_G(A^\infty)$ such that

¹A proof of these two citations also can be found in the appendix.

- $w \in L(w) \subseteq \llbracket w \rrbracket_\varphi$.
- the number of such languages is bounded by some function in $|A|$ and $|M|$.
- every word in $L(w)$ starts with the same letter.

In particular, we want to saturate $\llbracket w \rrbracket_\varphi$ by sets in $\text{SD}_G(A^\infty)$. The construction of the set $L(w)$ is by induction on $(|M|, |A|)$ with lexicographic order.

If $w = 1$, then we set $L(w) = \{1\}$. This concludes the induction base $|A| = 0$. Let us consider the case that $\varphi(A^*)$ is a group, that is, a divisor of G . Consider the prefix code $K = A$ of synchronization delay 1 and the homomorphism $\gamma = \varphi$. Note that since $\{a\} \in \text{SD}_G(A^\infty)$ and $\text{SD}_G(A^\infty)$ is closed under union, every subset of K is in $\text{SD}_G(A^\infty)$. In particular, $K \cap \gamma^{-1}(g) \in \text{SD}_G(A^\infty)$ for all $g \in \varphi(A^*)$. This shows $\gamma^{-1}(g) = \varphi^{-1}(g) \in \text{SD}_G(A^*)$ for all $g \in \varphi(A^*)$ by Lemma 2 and Lemma 3. If $w = av \in aA^*$ for some $a \in A$, then set $L(w) = a\varphi^{-1}(\varphi(v))$. It is clear that $w \in L(w) \subseteq \llbracket w \rrbracket_\varphi$ and $L(w) \in \text{SD}_G(A^\infty)$ by the above. If $w \in aA^\omega$, then we obtain $w \in a\varphi^{-1}(m)\varphi^{-1}(1)^\omega$ for some $m \in M$ by Ramsey's theorem. The idempotent in this decomposition must be 1 since $\varphi(A^*)$ is a group. Thus, we may set $L(w) = a\varphi^{-1}(m)\varphi^{-1}(1)^\omega$. Note that by the definition of \sim_φ , the inclusion $L(w) \subseteq \llbracket w \rrbracket_\varphi$ holds. In particular, these cases include the induction base $|M| = 1$.

In the following we assume that $\varphi(A^*)$ is not a group and therefore there exists a letter $c \in A$ such that $\varphi(c)$ is not a unit. Fix this letter $c \in A$ and set $B = A \setminus \{c\}$. If $w \in B^\infty$, the set $L(w)$ exists by induction. Let $w = uv$ with $u \in B^*$ and $v \in cA^\infty$. By induction we obtain $L(u) \in \text{SD}_G(B^\infty) \subseteq \text{SD}_G(A^\infty)$ and it remains to show $L(v) \in \text{SD}_G(A^\infty)$. Note that the product $L(w) = L(u) \cdot L(v)$ is unambiguous. From now on we may assume $w \in cA^\infty$. Let us first consider the case $w = uv$ with $u \in c(B^*c)^*$ and $v \in B^\infty$, i.e., there are only finitely many occurrences of the letter c in w . By induction, there exists $L(v) \in \text{SD}_G(B^\infty) \subseteq \text{SD}_G(A^\infty)$ and by setting $L(w) = L(u) \cdot L(v)$ it remains to construct $L(u)$.

Consider the alphabet $T = \varphi(B^*) = \{\varphi(u) \mid u \in B^*\}$. Let M_c be the local divisor of M at $\varphi(c)$. Since M_c is a divisor of M , every subgroup of M_c is a divisor of G . Consider the homomorphism $\psi : T^* \rightarrow M_c$ given by $\psi(\varphi(u)) = \varphi(cuc)$ and the substitution $\sigma : (B^*c)^\infty \rightarrow T^\infty$ with $\sigma(u_1cu_2c\dots) = \varphi(u_1)\varphi(u_2)\dots$. Note that

$$\begin{aligned} \psi(\sigma(u_1cu_2c\dots u_nc)) &= \psi(\varphi(u_1)\varphi(u_2)\dots\varphi(u_n)) = \varphi(cu_1c) \circ \varphi(cu_2c) \circ \dots \circ \varphi(cu_nc) \\ &= \varphi(cu_1cu_2c\dots cu_nc) \end{aligned}$$

and thus $\varphi^{-1}(m) \cap c(B^*c)^* = c\sigma^{-1}(\psi^{-1}(m))$. By induction on the monoid size, since $|M_c| < |M|$, there exists a language $L(\sigma(u')) \in \text{SD}_G(T^\infty)$ for all $u' \in (B^*c)^*$. We show $\sigma^{-1}(K) \in \text{SD}_G(A^\infty)$ for all $K \in \text{SD}_G(T^\infty)$ inductively on the definition of SD_G . Then we can set $L(u) = c\sigma^{-1}(L(\sigma(u')))$ for $u = cu'$ and have completed the case of finitely many c 's.

For $K = \emptyset$, we obtain $\sigma^{-1}(K) = \emptyset \in \text{SD}_G(A^\infty)$. Furthermore,

$$\sigma^{-1}(t) = \bigcup_{v \in B^*, t = \varphi(v)} L(v)c \in \text{SD}_G(A^\infty).$$

Let $L, K \in \text{SD}_G(T^\infty)$. A basic result from set theory yields $\sigma^{-1}(L \cup K) = \sigma^{-1}(L) \cup \sigma^{-1}(K)$. Let $\sigma(v) = w_1 w_2$ for some $v \in (B^*c)^*$. Since B^*c is a prefix code, there exists a unique factorization $v = v_1 v_2$ with $v_1, v_2 \in (B^*c)^*$ such that $\sigma(v_1) = w_1$ and $\sigma(v_2) = w_2$. Thus, we conclude $\sigma^{-1}(K \cdot L) = \sigma^{-1}(K) \cdot \sigma^{-1}(L)$. Let now $K \in \text{SD}_G(T^\infty)$ be a prefix code of synchronization delay d . We first show that $\sigma^{-1}(K)$ is a prefix code of bounded synchronization delay. Let $u, uv \in \sigma^{-1}(K)$, then $\sigma(u), \sigma(uv) = \sigma(u)\sigma(v) \in K$ and therefore $\sigma(v) = 1$. This implies $v = 1$ and $\sigma^{-1}(K)$ is a prefix code. We prove that $\sigma^{-1}(K)$ has synchronization delay $d + 1$. The incrementation of the synchronization delay by one comes from the fact that B^*c is not a suffix code, and thus we need another word in B^*c to pose as a left marker. Consider $uvw \in \sigma^{-1}(K)^*$ with $v \in \sigma^{-1}(K)^{d+1}$ and factorize $v = v_1 c v_2$ with $v_2 \in \sigma^{-1}(K)^d = \sigma^{-1}(K^d)$. Then $\sigma(uvw) = \sigma(uv_1 c) \sigma(v_2) \sigma(w)$, and by $\sigma(v_2) \in K^d$ this implies $\sigma(uv) = \sigma(uv_1 c) \sigma(v_2) \in K^*$. Thus, $uv \in \sigma^{-1}(K)^*$. Let $\gamma : K^* \rightarrow G$ be some homomorphism and $K_g = K \cap \gamma^{-1}(g) \in \text{SD}_G(T^\infty)$ for all $g \in G$. Inductively, $\sigma^{-1}(K_g) \in \text{SD}_G(A^\infty)$ and $\sigma^{-1}(K) = \bigcup \sigma^{-1}(K_g)$. Let $\gamma' : \sigma^{-1}(K)^* \rightarrow G$ be induced by $\gamma'(u) = \gamma(\sigma(u))$. By definition of $\text{SD}_G(A^\infty)$ we obtain $\gamma'^{-1}(1) \in \text{SD}_G(A^\infty)$. However, $u_1 \cdots u_n \in \sigma^{-1}(\gamma^{-1}(1))$ if and only if $\gamma(\sigma(u_1 \cdots u_n)) = 1$. Furthermore, note that $\gamma(\sigma(u_1 \cdots u_n)) = \gamma(\sigma(u_1)) \cdots \gamma(\sigma(u_n)) = \gamma'(u_1) \cdots \gamma'(u_n) = \gamma'(u_1 \cdots u_n)$. Thus, we obtain $\sigma^{-1}(\gamma^{-1}(1)) = \gamma'^{-1}(1) \in \text{SD}_G(A^\infty)$ and $\sigma^{-1}(\gamma^{-1}(1)^\omega) = \gamma'^{-1}(1)^\omega \in \text{SD}_G(A^\infty)$.

The last case of the proof is that w contains infinitely many c 's, that is, $w = cv$ with $v \in (B^*c)^\omega$. By induction, we know that $\sigma(v) \in L_T \cdot \gamma_T^{-1}(1)^\omega \subseteq \llbracket \sigma(v) \rrbracket_\psi$ for some $L_T \in \text{SD}_G(T^*)$ and $\gamma_T : K_T^* \rightarrow G$ for some prefix code $K_T \in \text{SD}_G(T^*)$ of bounded synchronization delay with $\gamma_T^{-1}(g) \cap K_T \in \text{SD}_G(T^*)$. By the calculation above, there exists a $\gamma : K^* \rightarrow G$ with the usual properties such that $\gamma^{-1}(1) = \sigma^{-1}(\gamma_T^{-1}(1))$. Let $L = \sigma^{-1}(L_T)$ and set $L(w) = cL\gamma^{-1}(1)^\omega$. It remains to show that $cL\gamma^{-1}(1)^\omega \subseteq \llbracket w \rrbracket_\varphi$. Let $cu \in cL\gamma^{-1}(1)^\omega$, then $\sigma(u) \in \llbracket \sigma(v) \rrbracket_\psi$, that is $\sigma(u) \approx_\psi \sigma(v)$. Since \approx_ψ is the transitive closure of \sim_ψ , we show that $\sigma(u) \sim_\psi \sigma(v)$ implies $cu \approx_\varphi cv$ for all $u, v \in (B^*c)^\omega$ which concludes the proof. Now, let $\sigma(u) = \sigma(u_1 c) \sigma(u_2 c) \cdots$ and $\sigma(v) = \sigma(v_1 c) \sigma(v_2 c) \cdots$ such that $\psi(\sigma(u_i c)) = \psi(\sigma(v_i c))$. As observed above, this implies $\varphi(cu_i c) = \varphi(cv_i c)$. Thus,

$$\begin{aligned} cu &= (cu_1 c) u_2 (cu_3 c) u_4 (c \cdots \sim_\varphi (cv_1 c) u_2 (cv_3 c) u_4 (c \cdots \\ &= cv_1 (cu_2 c) v_3 (cu_4 c) \cdots \sim_\varphi cv_1 (cv_2 c) v_3 (cv_4 c) \cdots \\ &= cv. \end{aligned}$$

This implies the existence of sets $L(w) \in \text{SD}_G(A^\infty)$ with $w \in L(w) \subseteq \llbracket w \rrbracket_\varphi$ in the case of infinitely many c 's. \square

5. Rees extension monoids

In this section we prove the direction $2 \iff 3$. We need the fact that every group contained in $\text{Rees}(N, M, \rho)$ is contained in N or in M .

Lemma 12 ([1]). *Let G be a group in $\text{Rees}(N, M, \rho)$, then there exists an embedding of G into N or into M .*

Thus, Lemma 12 implies $\text{LocRees}(\mathbf{H}) \subseteq \text{Rees}(\mathbf{H}) \subseteq \text{Rees}(\overline{\mathbf{H}}) \subseteq \overline{\mathbf{H}}$ for any group variety \mathbf{H} , which is $3 \implies 2$. We want to prove equality, that is, every monoid which contains only groups in \mathbf{H} is a divisor of an iterated Rees extension of groups in \mathbf{H} . However, we are able to prove a stronger statement using only local Rees extensions.

Proposition 13. *Given M , we can construct a sequence of monoids $M_1, \dots, M_k = M$ with $k \leq 2^{|M|} - 1$ such that for each $1 \leq j \leq k$ we have for M_j one of the following:*

- M_j is a group which is a divisor of M .
- M_j is a divisor of a local Rees product of some M_i and a local divisor M_ℓ of M_j with $i, \ell < j$.

Proof. We proof the statement with induction on $|M|$. If M is a group, we set $M_1 = M$. This includes the base case $|M| = 1$. If M is not a group, we may choose a minimal generating set of M . Let c be a nonunit of this generating set, then there exists a proper submonoid N of M such that N and c generate M . Since c is not a unit, the local divisor M_c is smaller than M , that is, $|M_c| < |M|$. By induction, there exist sequences $M'_1, \dots, M'_{k'} = N$ and $M''_1, \dots, M''_{k''} = M_c$ with $k', k'' \leq 2^{|M|-1} - 1$. We show that M is a homomorphic image of the local Rees product $\text{LocRees}(N, M_c)$. Let $\varphi : \text{LocRees}(N, M_c) \rightarrow M$ be the mapping given by $\varphi(n) = n$ for $n \in N$ and $\varphi(u, x, v) = u x v$ for $(u, x, v) \in N \times M_c \times N$. Since

$$\begin{aligned} \varphi((u, x, v)(s, y, t)) &= \varphi(u, x \circ c v s c \circ y, t) = \varphi(u, x v s y, t) \\ &= (u x v)(s y t) = \varphi(u, x, v) \varphi(s, y, t), \end{aligned}$$

φ is a homomorphism. Obviously, $M = N \cup N M_c N$ and thus φ is surjective.

Setting $M_i = M'_i$ for $1 \leq i \leq k'$, $M_{i+k'} = M''_i$ for $1 \leq i \leq k''$ and $M_{k'+k''+1} = M$ leads to such a sequence for M as M is a divisor of the local Rees product of $M_{k'} = N$ and $M_{k'+k''} = M_c$. Since $k' + k'' + 1 \leq 2 \cdot (2^{|M|-1} - 1) + 1 = 2^{|M|} - 1$, the bound on k holds. \square

The inclusion $\overline{\mathbf{H}} \subseteq \text{LocRees}(\mathbf{H})$ is immediate from Proposition 13, which is $2 \implies 3$. In particular, every monoid in $\overline{\mathbf{H}}$ is a divisor of an iterated Rees product of groups in \mathbf{H} by Lemma 1. We can draw the decomposition as a tree based on the decomposition of M in submonoids and local divisors. We do not describe this formally but content ourselves to give an example.

Example 14. *Let M be the monoid generated by $\{a, b, \delta, \sigma\}$ with the relations $a^2 = b^2 = ab = ba = 0$, $a\delta = a$, $\delta\sigma = \sigma\delta^2$, $\delta^3 = 1$, $\sigma^2 = 1$ and $d\delta = \delta d$, $d\sigma = \sigma d$ with $d \in \{a, b\}$. The subgroup generated by δ and σ is the symmetric group \mathfrak{S}_3 ; it is solvable but not Abelian. The monoid M is syntactic for the language L which is a union of L_a and L_b . The language L_a is the set of all words uav with $uv \in \{\delta, \sigma\}^*$ and the sign of the permutation uv evaluates to -1 . The language L_b is the set of all words ubv with $uv \in \{\delta, \sigma\}^*$ and uv evaluates in \mathfrak{S}_3 to δ . The decomposition in Rees products from Proposition 13 is depicted in Figure 1. Here $M[a, \sigma, \delta]$ denotes the submonoid generated by $\{a, \sigma, \delta\}$. In particular, this yields $M \preceq \text{Rees}(\text{Rees}(\mathfrak{S}_3, \mathbb{Z}/2\mathbb{Z}, \rho_1), \text{Rees}(\mathfrak{S}_3, \{1\}, \rho_2), \rho_3)$ for some ρ_1, ρ_2, ρ_3 by Lemma 1.*

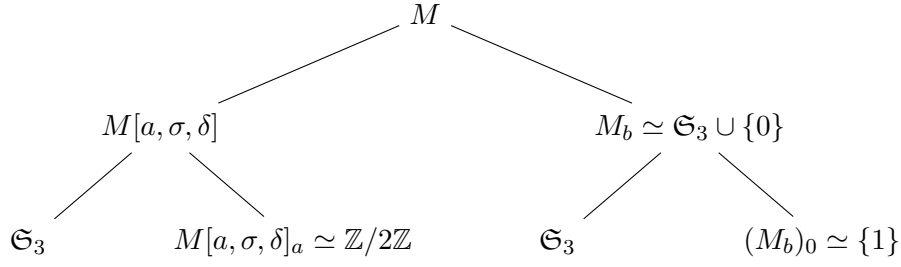


Figure 1: Decomposition tree of the monoid in Example 14.

6. Applications

An application of Proposition 13 is the solution to an open question of Almeida and Klíma. Let \mathbf{U} and \mathbf{V} be varieties. Let $\text{Rees}(\mathbf{U}, \mathbf{V})$ be the variety generated by $\text{Rees}(N, M, \rho)$ for $N \in \mathbf{U}$ and $M \in \mathbf{V}$. Note that in general $\text{Rees}(\mathbf{V}) \neq \text{Rees}(\mathbf{V}, \mathbf{V})$. However $\text{Rees}(\mathbf{V})$ can be defined as the limit of this operation. Let $\mathbf{V}_i = \text{Rees}(\mathbf{V}_{i-1}, \mathbf{V}_{i-1})$ and $\mathbf{V}_0 = \mathbf{V}$, then

$$\text{Rees}(\mathbf{V}) = \bigcup_{i \in \mathbb{N}} \mathbf{V}_i.$$

The variety $\text{Rees}(\mathbf{U}, \mathbf{V})$ has recently been introduced by Almeida and Klíma under the name of *bullet operation* [1]. They defined a variety \mathbf{V} to be *bullet idempotent* if $\mathbf{V} = \text{Rees}(\mathbf{V}, \mathbf{V})$ and posed the open question whether there are varieties apart from $\overline{\mathbf{H}}$ which are bullet idempotent. Using our decomposition above, we prove that the answer to this question is no.

Theorem 15. *Let \mathbf{V} be a bullet idempotent variety and let $\mathbf{H} = \mathbf{V} \cap \mathbf{G}$, then $\mathbf{V} = \overline{\mathbf{H}}$.*

Proof. Since $\overline{\mathbf{H}}$ is the maximal variety with $\overline{\mathbf{H}} \cap \mathbf{G} = \mathbf{H}$, we have $\mathbf{V} \subseteq \overline{\mathbf{H}}$. Let $M \in \overline{\mathbf{H}}$. Inductively, we may assume that every proper divisor of M is in \mathbf{V} . If M is a group, then $M \in \mathbf{H}$ and thus $M \in \mathbf{V}$. Thus, there exists a nonunit element $c \in M$ and a proper submonoid N of M such that N and c generate M . By the calculation in the proof of Proposition 13, M is a divisor of $\text{LocRees}(N, M_c)$, and since $N, M_c \in \mathbf{V}$ and $\mathbf{V} = \text{Rees}(\mathbf{V}, \mathbf{V})$ we obtain $M \in \mathbf{V}$. \square

Let $(\text{FO} + \text{MOD}_q)[<]$ be the fragment of first-order sentences which only use first-order quantifiers, modular quantifiers of modulus q and the predicate $<$. Then the following theorem holds.

Corollary 16. $(\text{FO} + \text{MOD}_q)[<](A^\infty) = \text{SD}_{\text{Sol}_q}(A^\infty)$

Proof. By [18], see also [17] for a complete treatise, $(\text{FO} + \text{MOD}_q)[<]$ describes the family of all regular languages such that every group in the syntactic monoid is a solvable group of cardinality dividing a power of q , that is the languages in Sol_q . Theorem 4 then implies the stated equality. \square

	$\overline{1}$	\overline{Ab}	\overline{Sol}	\overline{Sol}_q	\overline{H}
finite words	[15]	[14]	[16],new	[16],new	new, unless $H \subseteq Ab$
ω -words	[4]	new	new	new	new, unless $H = 1$

Table 1: Overview of existing and new language characterizations of \overline{H} .

The same language class has been described by Straubing with another operation, counting how many prefixes are in a given language, which resembles more closely the counting modulo q [16].

7. Summary

Our main theorem Theorem 4 states $\overline{H}(A^\infty) = SD_H(A^\infty)$. An overview over the contributions for \overline{H} is given in Figure 1. As a byproduct we were able to give a simple decomposition of the monoids in \overline{H} as local Rees products and groups in H , using only exponentially many operations.

References

- [1] J. Almeida and O. Klíma. On the irreducibility of pseudovarieties of semigroups. *Journal of Pure and Applied Algebra*, 220(4):1517 – 1524, 2016.
- [2] Jean-Camille Birget and John L. Rhodes. Almost finite expansions of arbitrary semigroups. *Journal of Pure and Applied Algebra*, 32(3):239–287, 1984.
- [3] Jean-Camille Birget and John L. Rhodes. Group theory via global semigroup theory. *Journal of Algebra*, 120(2):284 – 300, 1989.
- [4] Volker Diekert and Manfred Kufleitner. Omega-rational expressions with bounded synchronization delay. *Theory Comput. Syst.*, 56:686–696, 2015.
- [5] Volker Diekert and Manfred Kufleitner. A survey on the local divisor technique. *Theoretical Computer Science*, 610:13–23, 2015.
- [6] Volker Diekert, Manfred Kufleitner, and Pascal Weil. Star-free languages are Church-Rosser congruential. *Theoretical Computer Science*, 454:129–135, 2012.
- [7] Volker Diekert and Grzegorz Rozenberg, editors. *The Book of Traces*. World Scientific, Singapore, 1995.
- [8] Samuel Eilenberg. *Automata, Languages, and Machines*, volume B. Academic Press, New York and London, 1976.
- [9] Dominique Perrin and Jean-Éric Pin. *Infinite words*, volume 141 of *Pure and Applied Mathematics*. Elsevier, Amsterdam, 2004.

- [10] Jean-Éric Pin. *Varieties of Formal Languages*. North Oxford Academic, London, 1986.
- [11] John Rhodes and Dennis Allen. Synthesis of the classical and modern theory of finite semigroups. *Advances in Mathematics*, 11(2):238 – 266, 1973.
- [12] John L. Rhodes and Benjamin Steinberg. *The \mathfrak{q} -theory of finite semigroups*. Springer Monographs in Mathematics. Springer, 2009.
- [13] Marcel-Paul Schützenberger. On finite monoids having only trivial subgroups. *Information and Control*, 8:190–194, 1965.
- [14] Marcel-Paul Schützenberger. Sur les monoides finis dont les groupes sont commutatifs. *Rev. Française Automat. Informat. Recherche Opérationnelle Sér. Rouge*, 8(R-1):55–61, 1974.
- [15] Marcel-Paul Schützenberger. Sur certaines opérations de fermeture dans les langages rationnels. In *Symposia Mathematica, Vol. XV (Convegno di Informatica Teorica, INDAM, Roma, 1973)*, pages 245–253. Academic Press, 1975.
- [16] Howard Straubing. Families of recognizable sets corresponding to certain varieties of finite monoids. *Journal of Pure and Applied Algebra*, 15(3):305–318, 1979.
- [17] Howard Straubing. *Finite Automata, Formal Logic, and Circuit Complexity*. Birkhäuser, Boston, Basel and Berlin, 1994.
- [18] Howard Straubing, Denis Thérien, and Wolfgang Thomas. Regular languages defined with generalized quantifiers. *Inform. and Comput.*, 118(2):289–301, 1995.
- [19] Wolfgang Thomas. Automata on infinite objects. In J. van Leeuwen, editor, *Handbook of Theoretical Computer Science*, chapter 4, pages 133–191. Elsevier Science Publishers B. V., 1990.

A. Missing proofs

All missing proofs can easily be deduced from the existing literature; and pointers have been given in previous sections. However, in order to keep the paper self-contained we reproduce them in our notation. We first give a proof of Proposition 6. The statement has been proved by Schützenberger. We give a detailed proof following [14] loosely. We assume the reader to be familiar with basic concepts of formal language theory, such as deterministic finite automata and remind the classic theorem that the transformation monoid of a minimal deterministic finite automaton of a language is isomorphic to its syntactic monoid.

Proof of Proposition 6. Note that K^* is regular because $K = \bigcup \{K_g \mid g \in G\}$ is regular. Without restriction we may assume $K \neq \emptyset$ and we let d be the synchronization delay of K . If p denotes a state in some deterministic finite automaton (DFA) and if $u \in A^*$ is a word, then we write $p \mapsto p \cdot u$ to indicate that reading u transforms p into the state $p \cdot u$. For $g \in G$ let Q_g be the state set of the minimal automaton for K_g and q_g the corresponding initial state. Let Q be the direct product of sets Q_g with initial state $q_0 = \prod \{q_g \mid g \in G\}$. The product automaton allows to assign to each language K_g a subset $F_g \subseteq Q$ such that the DFA (Q, A, \cdot, q_0, F_g) accepts K_g . Since $K_g \cap K_h = \emptyset$ for $g \neq h$ we have $F_g \cap F_h = \emptyset$ for $g \neq h$. It is also clear that $\prod_{g \in G} \text{Synt}(K_g)$ acts on Q .

By F we denote union $\bigcup \{F_g \mid g \in G\}$. We merge the subset $\{p \in Q \mid p \cdot A^* \cap F = \emptyset\}$ into a single sink state \perp . Since K is a prefix code, there is no word $u \in A^+$ such that $p \cdot u \in F$ for any $p \in F$. Thus, $p \cdot u = \perp$ for every $p \in F$ and $u \in A^+$. Moreover, without restriction we may assume that every state is reachable from the initial state q_0 and by slight abuse of language, the new state space is still called Q . The image of A^* in the transformation monoid Q^Q which is induced by $\sigma_u : Q \rightarrow Q, p \mapsto p \cdot u$ defines a monoid S , the transition monoid of Q , and S becomes a divisor of $\prod_{g \in G} \text{Synt}(K_g)$. It is therefore enough to show that every subgroup in the syntactic monoid $\text{Synt}(\gamma^{-1}(1))$ is either a divisor of G or a divisor of S . For later use we denote by $\sigma : A^* \rightarrow S$ the homomorphism which maps u to σ_u .

Next, consider the product set $\tilde{Q} = G \times (Q \setminus F)$. We view \tilde{Q} as a state space of an automaton accepting $\gamma^{-1}(1)$ as follows.

$$(g, q) \cdot a = \begin{cases} (g, q \cdot a) & \text{if } q \cdot a \in Q \setminus F \\ (gh, q_1) & \text{if } q \cdot a \in F_h \end{cases}$$

Note that the transition function is well-defined since, as mentioned above, $F_g \cap F_h = \emptyset$ for $g \neq h$. The construction defines a homomorphism $\mu : A^* \rightarrow \tilde{Q}^{\tilde{Q}}$. We let $M = \mu(A^*)$. It is the corresponding transition monoid for \tilde{Q} . Moreover, letting $(1, q_1) \in \tilde{Q}$ be the only final state, the resulting DFA accepts $\gamma^{-1}(1)$ as a subset of A^* . To see this observe that every word $u \in \gamma^{-1}(1)^*$ belongs to $K^* \subseteq A^*$. Moreover, u admits a unique factorization $u = u_1 \cdots u_k$ such that for all i we have $q_0 \cdot u_i \in F_{g_i}$ for $g_i = \gamma(u_i)$ and $1 = g_1 \cdots g_k$. Since the DFA accepts $\gamma^{-1}(1)$, it is enough to show that every subgroup of M is either a subgroup of G or a divisor of S .

Let H be a subgroup of M . Then H contains a unique idempotent $e \in M$ which is the neutral element in H . In particular, $H = eHe$. Let $\mathcal{H} = \mu^{-1}(H)$. It is a nonempty subsemigroup of A^* . The group H does not act as a group on \tilde{Q} , because there might be states (g, p) such that $(g, p) \neq (g, p) \cdot e$. However, it acts faithfully on $\tilde{Q}_e = \tilde{Q} \cdot e$. Indeed, if $h \neq h'$ in H , then there are states $(g, p) \cdot h \neq (g, p) \cdot h'$. Since $h = ehe$ and $h' = eh'e$, we have $(g, p) \cdot e \in \tilde{Q}_e$, $(g, p) \cdot eh \neq (g, p) \cdot eh'$, and $(g, p) \cdot eh, (g, p) \cdot eh' \in \tilde{Q}_e$. We distinguish two cases.

Case 1. There is a state $(g, p) \in \tilde{Q}_e$ such that there is a word $uv \in \mathcal{H}$ where $p \cdot u \in F$. For $w = (uv)^{|H|}$ we have $\mu(w) = e$ and w factorizes as $w = uw'x$ such that $w' \in K^*$ and $q_0 \cdot x = p$. It follows $xu \in K$. Letting $y = wuw'$ we have $yx = w^2 \in \mathcal{H}$ with $\mu(yx) = e$ and hence, $(g, q_0) \cdot x = (g, p)$ implies $(g, p) \cdot y = (g, q_0)$.

The element $\mu(xy)$ is idempotent in M . Indeed, calculating in M we have:

$$(xy)^2 = xwu w' \cdot xwu w' = xw^3 w' = xwu w' = xy.$$

The subsemigroup xHy contains the idempotent xy and $f \mapsto xfy$ defines a homomorphism of H onto the group H' and its inverse is given $xfy \mapsto yxfyx = f$. As H and H' are isomorphic, we start all over with the idempotent $e' = \mu(xy)$, the group H' , and its inverse image \mathcal{H}' instead of e, H, \mathcal{H} .

In order to simplify the notation we rename e', H', \mathcal{H}' as e, H, \mathcal{H} . The difference is that, now, we have $(g, q_0) \cdot e = (g, q_0)$ and $\mu(xy) = e$ with $xy \in K^+$. Consider $(g, q) \in \tilde{Q}_e$ such that $q \neq \perp$ and hence, q is not the sink state of Q . Then there exist words $u, v \in A^*$ such that $q_0 \cdot u = q$ and $q \cdot v \in F$. Since $(g, q) = (g, q_0) \cdot u \in \tilde{Q}_e$, we obtain $(g, q_0) \cdot u(xy)^d v = (g, q) \cdot v = (g', q_0)$ for some $g' \in G$. Consequently, $u(xy)^d v \in K^*$ and, by synchronization delay, we obtain $u(xy)^d \in K^*$. In particular, $(g, q_0) \cdot u(xy)^d = (g, q_0)$. Thus, $(g, q) = (g, q_0) \cdot (xy)^d = (g, q_0)u(xy)^d = (g, q_0)$ and therefore, $q = q_0$. Thus,

$$\tilde{Q}_e \subseteq \{(g, q_0) \mid g \in G\} \cup \{(g, \perp) \mid g \in G\}.$$

This implies $\mathcal{H} \subseteq K^*$ by the definition of the automaton. (The group H acts trivially on $\{(g, \perp) \mid g \in G\}$ and this part is irrelevant in the following.)

Consider the mapping $\pi : H \rightarrow G$ given by $\pi(\mu(u)) = \gamma(u)$ for $u \in \mathcal{H}$. This mapping is well-defined, since $(g, q_0) \cdot \mu(u) = (g \cdot \gamma(u), q_0)$ for some $(g, q_0) \in \tilde{Q}_e$. Thus, the homomorphism $\gamma : \mathcal{H} \rightarrow G$ factorizes as follows:

$$\gamma : \mathcal{H} \xrightarrow{\mu} H \xrightarrow{\pi} G.$$

Let us show that the homomorphism π is injective. We know that H acts faithfully on \tilde{Q}_e . Hence for $h \neq 1$ there is some $(g, q) \in \tilde{Q}_e$ such that $(g, q) \cdot h \neq (g, q)$. Thus, $(g, q) = (g, q_0)$ and therefore,

$$(g, q) \cdot h = (g\pi(h), q_0) \neq (g, q_0).$$

This shows, as desired, $\pi(h) \neq 1$ and H is a subgroup of G .

Case 2. For every state $(g, p) \in \tilde{Q}_e$ and every $uv \in \mathcal{H}$ we have $p \cdot u \notin F$. Thus, for all $(g, p) \in \tilde{Q}_e$ and all $u \in \mathcal{H}$ we have

$$(g, p) \cdot \mu(u) = (g, p \cdot u) = (g, p \cdot \sigma(u)).$$

This means that H acts faithfully on the following set

$$Q' = \left\{ p \in Q \mid (g, p) \in \tilde{Q}_e \right\}.$$

Let S' denote the submonoid $S' = \{s \in S \mid Q' \cdot s \subseteq Q'\}$, then $\sigma(\mathcal{H}) \subseteq S'$ and H becomes a quotient of S' and therefore, a divisor of S . This concludes the proof. \square

Next, we introduce a variant of Schützenberger products to give a short proof that the concatenation product of two languages does not introduce new groups, [13]. Let M be a finite monoid and $\varphi : A^* \rightarrow M$ be a homomorphism. We define the set

$$[w] = \{(\varphi(w_1), \varphi(w_2)) \in M \times M \mid w = w_1 w_2\}.$$

Further, we define the operations

$$\begin{aligned} u \cdot [w] &= \{(\varphi(u)m, n) \mid (m, n) \in [w]\} \\ [w] \cdot u &= \{(m, n\varphi(u)) \mid (m, n) \in [w]\}. \end{aligned}$$

One can check that $u \cdot [v] \cup [u] \cdot v = [uv]$. Our variant of the Schützenberger product is defined as the monoid

$$\tilde{M} = \{[w] \in M \times M \mid w \in A^*\}$$

equipped with the operation $[u][v] = [uv]$. This is well-defined since $[u] = [v]$ implies $\varphi(u) = \varphi(v)$. In fact, $\tilde{\varphi} : \tilde{M} \rightarrow M$ given by $\tilde{\varphi}([w]) = \varphi(w)$ is a homomorphism. It is fairly easy to see that \tilde{M} recognizes the concatenation product over A^∞ as well, see [7, Proposition 11.7.10].

Proposition 17. *Let $L \subseteq A^*$ and $K \subseteq A^\infty$ be languages recognized by $\varphi : A^* \rightarrow M$. Then $L \cdot K$ is recognized by the homomorphism $\psi : A^* \rightarrow \tilde{M}$ given by $\psi(w) = [w]$.*

Proof. Let $u = u_1 u_2 \in A^*$ such that $u_1 \in L$ and $u_2 \in K$ and consider some word $v \in A^*$ such that $\psi(u) = \psi(v)$. Since $(\varphi(u_1), \varphi(u_2)) \in [u] = [v]$, there exists a decomposition $v = v_1 v_2$ such that $(\varphi(u_1), \varphi(u_2)) = (\varphi(v_1), \varphi(v_2))$. Consequently, $v_1 \in L$ and $v_2 \in K$, i.e., $v \in L \cdot K$.

In the case of infinite words let $u = u_1 u_2 \dots \in L \cdot K$ and $v = v_1 v_2 \dots$ such that $\psi(u_i) = \psi(v_i)$ for all $i \in \mathbb{N}$, i.e., $u \sim_\psi v$. We may assume that $u_1 = u' u''$ such that $u' \in L$ and $u'' u_2 \dots \in K$. Again, there must exist a factorization $v_1 = v' v''$ such that $\varphi(u') = \varphi(v')$ and $\varphi(u'') = \varphi(v'')$. In particular, $v' \in L$. Since $\psi(u_i) = \psi(v_i)$ implies $\varphi(u_i) = \varphi(v_i)$, this yields $(u'' u_2) u_3 \dots \sim_\varphi (v'' v_2) v_3 \dots$ and therefore $v'' v_2 v_3 \dots \in K$. Thus, $v \in L \cdot K$, which completes the proof. \square

We show that every group contained in \tilde{M} is a group in M . The argument is a slight deviation of the original argument of Schützenberger and Petrone [13, Remark 2], in order to adapt to our variant of the Schützenberger product.

Proposition 18. *Let $\varphi : A^* \rightarrow M$ be a homomorphism and \tilde{M} be the corresponding Schützenberger product. Every group $G \subseteq \tilde{M}$ can be embedded into M .*

Proof. Let $[e]$ be the unit in G . Consider again the homomorphism $\tilde{\varphi} : \tilde{M} \rightarrow M$. Since G is finite, the set $N = \{[w] \in G \mid \tilde{\varphi}([w]) = \varphi(w) = \varphi(e) = \tilde{\varphi}([e])\}$ is a subgroup of N . In fact, N is normal and G/N is isomorphic to $\tilde{\varphi}(G)$, which is a group in M . Thus, it remains to show $N = \{[e]\}$, i.e., $\tilde{\varphi}$ is injective on G .

Let $[s] \in N$ be an arbitrary element and $[t] \in N$ be its inverse. Then, the following equations holds:

$$\bullet [e]^2 = [e] \qquad \bullet [e] = [s][t] \qquad \bullet [s] = [e][s][e]$$

By the first equation we have $[e] = e[e] \cup [e]e$.

By the second equation and $\varphi(s) = \varphi(t) = \varphi(e)$, it holds $[e] = s[t] \cup [s]t = e[t] \cup [s]e$. Since $e[e] \subseteq [e]$, we conclude $e[s]e \subseteq [e]$. Finally, using the third equation, we obtain

$$[s] = e([s][e]) \cup [e]se = e(s[e] \cup [s]e) \cup [e]e = e[e] \cup e[s]e \cup [e]e = [e] \cup e[s]e = [e].$$

□